利用分段直線迴歸模式

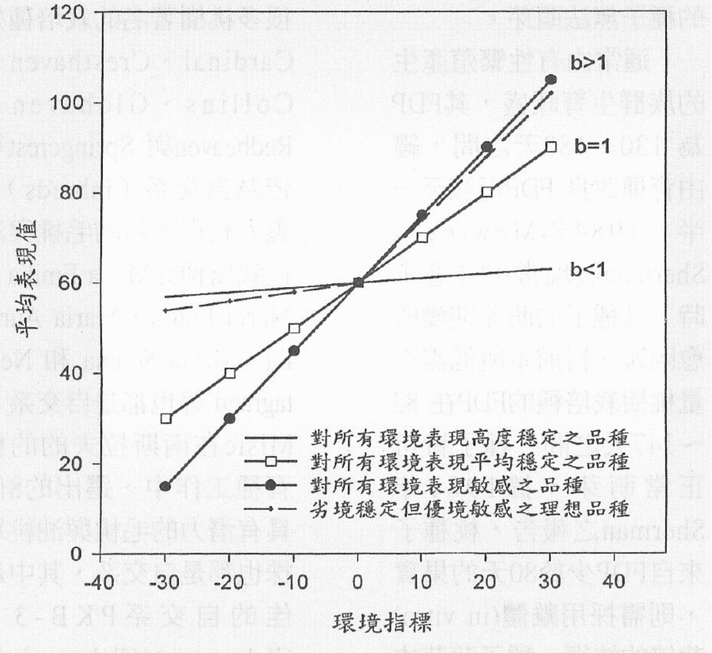
選拔高適應性作物品種

◎農試所農藝系　呂秀英

目前本省水稻及各種雜糧作物之育種實施程序中，均規定新育成品系在命名登記之前必須於各農業改良及試驗場所進行至少兩年的區域試驗，以選出相對於推廣品種之高產且穩定的品種。本省區域試驗一般採用的穩定性介量多以最受廣泛使用的直線迴歸分析法(Finaly and Wilkinson, 1963; Eberhart and Russell, 1966)為主。該法首先是以各環境之平均產量與整個試驗的總平均產量之偏差，作為該一試驗的環境效應估值，稱為「環境指標」，然後將各品種在各環境之產量隨這些指標作迴歸分析(regression analysis)，並以最小平方法（least square method）估計迴歸係數，利用這些迴歸係數估值（b）可以評價品種的穩定性：

當b＝1時，表示品種為符合農業觀念的穩定性，其會隨著環境改善而表現增進，但也不致因環境不利而表現太差；當b＞1時，為不穩定的品種 ；當b＝0時，則表示該品種不論環境如何變化，均能維持其一致的表現，屬生物觀念的穩定性，此多應用於病蟲害防治的品種選拔。

傳統直線迴歸穩定性分析法雖然具有計算簡單且容易解釋的優點，但是它的使用效益迄今仍備受爭議。除了其在統計數理



圖一、品種對環境表現不同穩定性的反應區別

及應用範圍上有所限制之外，引起最多的質疑是：該法真的能選出真正理想的品種嗎？因為一個真正符合農業觀念的理想品種，應該是在惡劣環境下仍能保持高產且穩定的表現，但在有利環境下可發揮其最大的產量增進能力，換言之，這種對惡劣環境鈍感而對有利環境敏感的品種，才具有最高的適應性，為農業育種選拔上最理想的基因型。由於這種理想的品種對環境優劣的表現互異，實際上是對所有環境的反應呈現一種非線性關係，如圖一所示。然而，傳統的直線迴歸穩定性分析法，是在假設所有品種對環境均具直線反應之前提下，將各品種對所有環境指標作迴歸分析，當品種在惡劣環境及有利環境的表現確實存有反向差異時，很有可能會產生互相抵銷的遮蓋效應，以致無法選出真正理想的品種。

對此，Verma and Chahal (1978)首先將環境指標分割為正負兩個子集合，然後再將各品種分別對該兩個環境子集合進行傳統

表一、傳統直線迴歸模式與分段直線迴歸模式之公式比較

	公　　式	穩定性係數
傳統直線	$Y_{ij} = m_i + b_i E_j$	b
分段直線	$Y_{ij} = m_i + b1_i E_j + b2_i (E_j - Ex) I_j$	$b_p = b_1$
	其中當 $E_j > E_x$ 時 $I_j = 1$，否則 $I_j = 0$	$b_f = b_1 + b_2$

註：Y_{ij}為第i品種在第j環境下的性狀平均值；m_i為第i品種在所有環境下的平均值；E_j為第環境之環境指標估計值；Ex為連接兩段迴歸直線的已知結點，此處設定為0，即以Ej之正負值來區分優劣環境。

直線迴歸分析，由兩條迴歸線所估得的兩個迴歸係數各表示品種在優劣環境下的穩定性係數。該法在觀念上相當簡單且實用，同時由於品種對優劣環境之穩定性評估結果間的相關性並不顯著，因此將所有環境指標分割成兩部分，再分別進行迴歸分析，其彼此間的評估結果將不會互相影響。但是直線迴歸法應用於穩定性分析上，有其根本之限制，即基因型及環境數不得太少，因此當有限的環境數再被分割成兩個子集合時，將勢必造成環境數會變得更少，而影響各子集合分別所估得之迴歸線斜率的顯著性，以致降低了穩定性分析的準確度。換言之，每一品種分別估算兩條迴歸線，需要更多的環境或重複數。事實上Verma and Chahal的方法，與假設兩段迴歸直線間之結點（knot）已知為0的分段直線迴歸模式（piece-wise linear regression model）之觀念完全相同，只不過後者是以含有兩個迴歸係數的單一個複迴歸方程式的形式來表達而已；經由分段直線迴歸模式所求配的參數估計值，理論上與Verma and Chahal法所得者相同，但多少可避免因為分開估算兩條迴歸線之情形下所導致之直線不易顯著的問題。傳統直線迴歸模式與分段直線迴歸模式之公式列如表一，由此可以比較出其間的差異性。

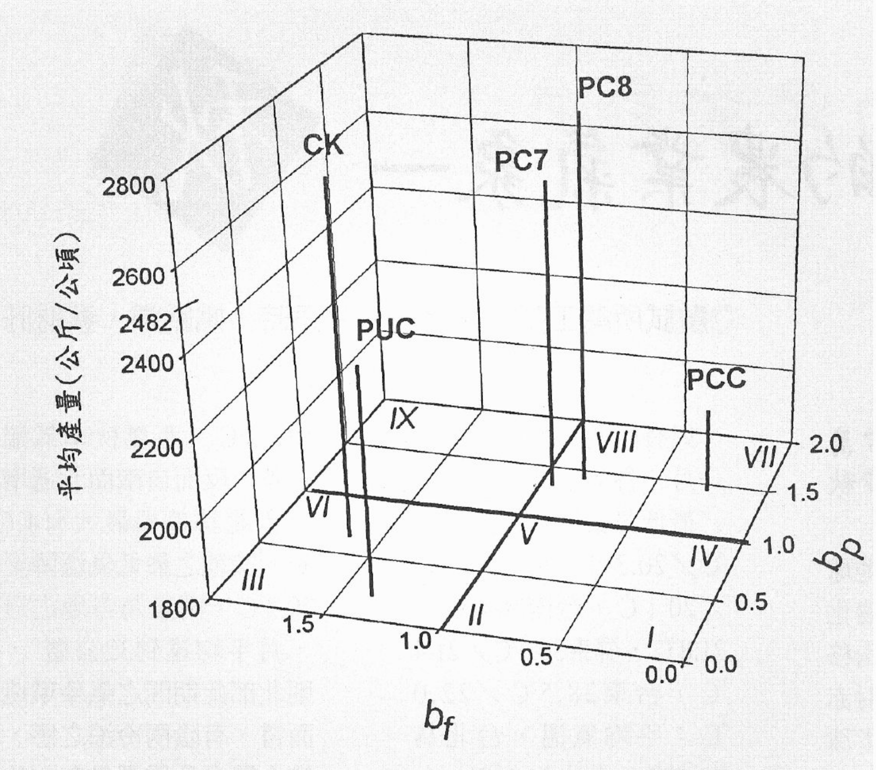
利用分段直線迴歸模

表二、根據分段直線迴歸法之劣境及優境穩定性係數估值進行基因型適應性之分類

類別	穩定性係數*		適應性	
	b_p	b_f	劣境	優境
I	< 1.0	< 1.0	高度穩定	高度穩定
II		= 1.0	高度穩定	平均穩定
III		> 1.0	高度穩定	敏感
IV	= 1.0	< 1.0	平均穩定	高度穩定
V		= 1.0	平均穩定	平均穩定
VI		> 1.0	平均穩定	敏感
VII	> 1.0	< 1.0	敏感	高度穩定
VIII		= 1.0	敏感	平均穩定
IX		> 1.0	敏感	敏感

* b_p ,b_f : 利用分段直線迴歸模式所估得之劣境及優境穩定性係數.

式進行穩定性分析，將產生兩個穩定性估計值：劣境穩定性係數（b_p）及優境穩定性係數（b_f）。根據b_p及b_f，基因型在不同環境下的相異表現有九種可能，如表二所示：其中第I類為無論環境如何變化均表現最為穩定的基因型，這種穩定性屬於生物觀念的穩定性，對農業上育成高產品種之改良幫助不大，但有利於病蟲害品種選拔；第II、III類基因型在惡劣環境下表現穩定，而在有利環境下則產量可獲增進，其增進效果完全決定於基因型平均值之表現，當平均值同時具有高產表現時為農業品種選拔上最理想的基因型，因為當環境改善的程度愈好，其表現也愈好，但在逆境時則仍能維持一定的水準，但當第II、III類基因型不具有高產表現時，表示其在有利環境下的增進效果不如其他基因型；反之，第VII類基因型在惡劣環境下表現敏感，但在環境變好時卻未見產量有明顯改善，此為最不好的基因型；第IX類為無論環境如何變化均最為敏感，是最不穩定的基因型；至於其他種類的基因型，其表現則介於上述各種類別之間。由於農業上多同時以穩定性估計值與產量平均值來作為選拔高產且穩定品種之依據，故將所估得的bp、bf及平均產量值三者之間的關係繪成圖示，則有利於基因型適應性的分類及篩選，此在參試品系數甚多的場合尤其有用。以五個玉米品系在六種環境之區域試驗資料的分段直線迴歸穩定性分析為例說明，結果如圖二所示，根據b_p及b_f兩座標軸可將所有品系區分為如表二所述之九種類別的適應性；各品系的產量平均值，則可由各品系垂直線的高度來進行比較。結果得知（圖二），PUC與對照品系的適應性屬於同一群（第III類），雖為具高適應性的品系，其對劣境鈍感而對優境則敏感，但PUC的產量平均值卻不如對照品系，也未高於產量總平均值(2482 kg ha^{-1})；PCC的適應性最差，其對劣境敏感而對優境鈍感（第VII類），且產量平均值也最低，為最不好的基因型；PC8及PC7



圖二、玉米五個品系之裂境穩定性係數（bp)、優境穩定性係數（bf）及平均產量間的關係

的適應性相近（第VIII類），雖然對環境改善能獲得產量增進（b_f近於1），但對惡劣環境也相當敏感（b_p＞1），不過因為PC8的產量優於對照品系，表示在本試驗環境範圍內PC8因處於有利環境所獲得的產量增進，足以彌補其在劣境下的產量損失；但仍然必須留意該品系往後可能受到不利環境所引起的低產風險。

Verma and Chahal 之分段直線迴歸模式，係針對傳統直線迴歸穩定性分析所提出的修正迴歸法，對評估廣大環境範圍之適應性，在觀念上頗為實用，尤其在基因型對環境呈非線性反應的場合。經由本人於1998年所進行之模擬試驗研究成果，也證實了傳統迴歸模式在評估作物對環境之非線性反應的不適用性，而利用分段直線迴歸模式可以有效地檢測出品系對優劣環境的反應差異。在應用分段直線迴歸模式於區域試驗之穩定性分析時，必須注意兩點：(一) 目前所提出的分段直線迴歸模式是假設所有品系均在環境指標數值等於0時從直線變成曲線，即分段迴歸的結點設定在0。事實上，各品系對環境反應的形式未必相同，即各品系的結點未必就會發生在同一點上，一旦如此，首先應該找出從直線變成曲線的結點是發生在哪一個環境範圍內，然後再利用分段直線迴歸模式。(二) 以栽培者的標準來說，田間研究中所謂的劣境，並不見得就是不利於作物生長的環境，因此在使用分段直線迴歸模式於穩定性分析時，除了利用環境指標之正負值將環境以武斷方式等分之外，或許也可憑著經驗或知識將環境分為劣、優兩類。